第1章

PCI Expressの基礎知識

プロトコル階層や物理層の基本がよく分かる

畑山 仁

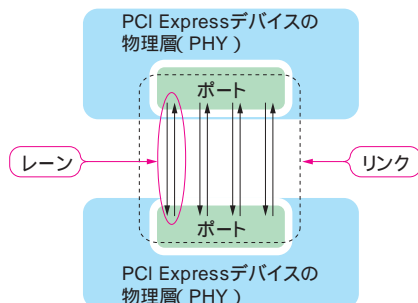
本稿ではPCI Expressを理解する上で必要な，プロトコルの各階層の概要や物理層の論理サブブロック，電気サブブロックの基本的な処理などの基礎知識を解説する．（編集部）

1 PCI Expressの基礎知識

● PCI Expressは帯域幅に応じてレーン数を選べる

PCI Expressの主な特徴を以下に示します．

- レーンあたり2.5Gbps，5Gbpsのデータ・レートが選べる
- 1，2，4，8，12，16，32レーンの帯域幅が選べる
- 送受信それぞれのシリアル伝送
- データの信頼性，電力管理，エラー・ロギング/レポートなどの機能
- レガシなPCIアーキテクチャをサポートすることでソフトウェア資産を継承
- ホット・プラグ/ホット・スワップ（活線挿抜）
- テストが容易である
- 普及のために，電気的な規格は，FR-4基板に対応

図1 PCI Expressのポート，レーン，リンクの関係
各レーンはそれぞれ送信用の差動信号と受信用の差動信号を持つ．レーンをまとめたものをリンクという．

本稿では，これらの特徴を理解するために必要なPCI Expressの基礎知識について解説します．

● 複数のレーンをまとめたものをリンクと呼ぶ

PCI Expressの各デバイスは，ほかのデバイスと接続するためのポートを備えています．ポートは，**図1**のような双方向通信を行うために，送受信の1組の差動ペアを単位とした「レーン」で構成されます（双対単方向伝送：デュアル・シンプレックス）．送信と受信が独立に同時にデータ転送できます．

さらにデータ帯域幅を上げるために，複数のレーンに拡張可能です．レーンをまとめたものをリンクと呼び，x1，x2，x4，x8，x12，x16，x32リンクが規格化されています．ここでxは「バイ」と呼び，x*N*リンクとは*N*組のレーンで構成されていることを意味します．

例えば現在，パソコン内部ではグラフィックス用にx16リンクが，外部I/O用に×1リンクが使われています．サーバではx8リンクが使用されています．なお，x2，x12，x32リンクはほとんど使われていません．

● ルート・コンプレックスの下にツリーを作る

PCI Expressのシステムを構成する要素として，**図2**のようにルート・コンプレックス，エンドポイント，スイッチ，ブリッジがあります．

1) ルート・コンプレックス

ルート・コンプレックスは，その名のように階層の根幹（Root）となるデバイスです．ルート・コンプレックスは一

KeyWord
PCI Express，トランザクション層，データ・リンク層，物理層，TLP，DLLP，レーン，リンク，スクランブル，8b/10b，Kコード

つ，あるいは複数のPCI Expressポートを持ちます．ホスト・ブリッジを内蔵し，CPUやメモリにも接続されます．

2）エンドポイント

I/OデバイスをPCI Expressではエンドポイントと呼びます．レガシ・エンドポイント，PCI Expressエンドポイント，ルート・コンプレックス・エンドポイントの3種類あります．

3）スイッチ

スイッチはPCI Expressポートを増やすためのデバイスです．

4）ブリッジ

ブリッジはプロトコル変換を行うデバイスです．特にPCI Expressでブリッジというと，PCI/PCI-Xを接続するためのデバイスを指すようです．

PCI Expressでは，これらのデバイスがPCIアーキテクチャとして，ルート・コンプレックスからのツリー構造をとります．なお，PCI Expressをベースとした ASI（Advanced Switch Interconnect）では，スター型やメッシュ型のトポロジも使えます．

● プロトコル階層に応じて役割がある

PCI Expressのプロトコルは図3のように階層化された構成をとります．トランザクション層，データ・リンク層，物理層で構成されます．上位層で生成されたパケットは，図4のように下位層に移るとともに，必要な情報が付加されます．受信側では逆の処理を行い，必要なデータを取り出します．

1）トランザクション層

トランザクション層は，トランザクション層パケット（TLP：Transaction Layer Packet）の構築と処理を行います．以下のようなTLPがあり，データの読み出し，書き込みなどのトランザクションのために使用されます．

- **メモリ・リクエスト**
  メモリに対する読み出し/書き込みを要求
- **I/Oリクエスト**
  I/Oに対する読み出し/書き込みを要求．
- **コンフィグレーション・リクエスト**
  コンフィグレーション空間に対する読み出し/書き込みを要求
- **コンプリッション**
  リクエスト・パケットに対する応答（読み出しの場合にはデータが含まれる）
- **メッセージ**
  割り込みやパワー・マネジメント・リクエストなど

TLP送信を調節するためにフロー制御（FC）を行うこともPCI Expressの特徴です．受信側のバッファの空きを確認してから，データの転送を開始します．そのため各デバイスは，TLPのためのFCクレジット・ステータスを，データ・リンク層を使用して，周期的に送信します．

図2　PCI Expressの構成要素
スイッチやエンドポイントなどがルート・コンプレックスからのツリー構造をとる．

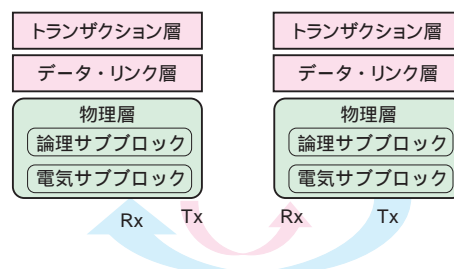
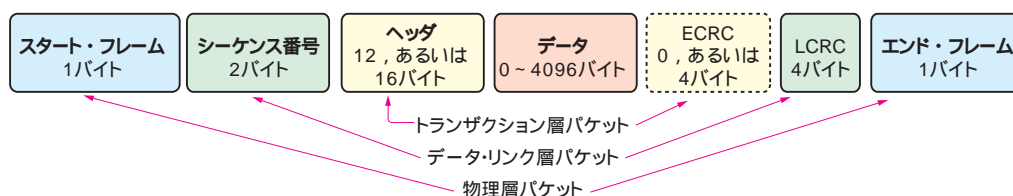



図3　PCI Expressのプロトコル階層
物理層の上にデータ・リンク層，トランザクション層が構成される．物理層は，論理サブブロックと電気サブブロックに分けられる．

図4
物理層で見たPCI Expressのパケット
下位の層に移るとともに必要な情報が付加される．

レスポンスを要求するすべてのリクエストは，レスポンスを待たずに要求を発行するスプリット・トランザクションです．各リクエスト・パケットは，レスポンス側が正確に要求元を対応付けられるように，固有の識別情報を与えます．

TLPは図5のようにヘッダから始まり，データ・ペイロード，場合によってTLPダイジェスト（ECRC ： End-to-end CRC）が続きます．12～16バイトのTLPヘッダの例を図6に示します．

2）データ・リンク層

データ・リンク層の主要な役割は，リンク上の二つのデバイス間でTLPを確実に交換するためのメカニズムを提供することです．この役割を果たすために，データ・リンク層はリンクを管理し，エラー検知およびエラー訂正によってデータの品質を維持します．

データ・リンク層の送信側は，トランザクション層によって組み立てられたTLPに対し，12ビットのTLPシーケンス番号と32ビットのCRC（LCRC）を計算・付加し，送信用物理層に渡します（図7）．受信側のデータ・リンク層は，受信したTLPの完全性をチェックし，トランザクション層にそれらを引き渡します．

データ・リンク層が情報を正確に受信できた場合には，ACKを返します．TLPエラーが検出され，受信を失敗した場合にはNAKを返し，TLPの再送を要求します．

データ・リンク層は，さらにリンク管理のために，図8に構造を示すデータ・リンク層パケット（DLLP ： Data Link Layer Packet）と呼ばれるパケットの送受信も行います．

DLLPには，

- ACK/NAK：アクノリッジ，再送要求
- Init FC1/Init FC2：トランザクション層で行うフロー

図5 TLPの構造
12～16バイトのヘッダに続いてデータが送信される．オプションで誤り検出用のECRCを付加できる．

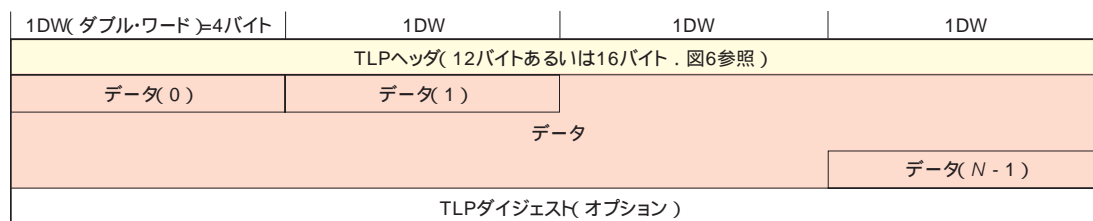



図6 TLPヘッダの例
64ビット・アドレスのメモリ・リクエストの場合の16バイトTLPヘッダの例を示す．

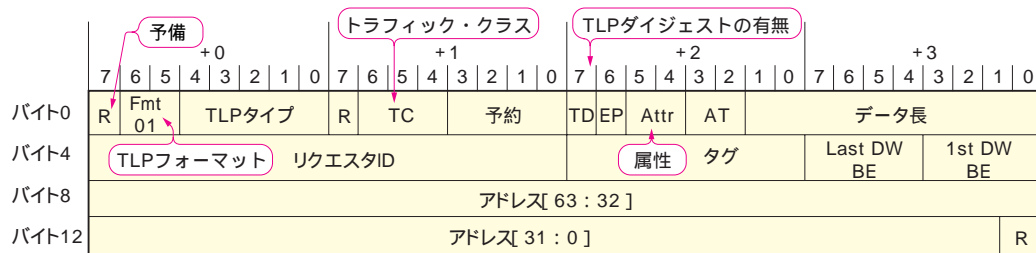



図7 データ・リンク層のTLP処理
データ・リンク層ではデータの品質保持のために，12ビットのTLPシーケンス番号および32ビットのCRC（LCRC）を計算し，付加する．

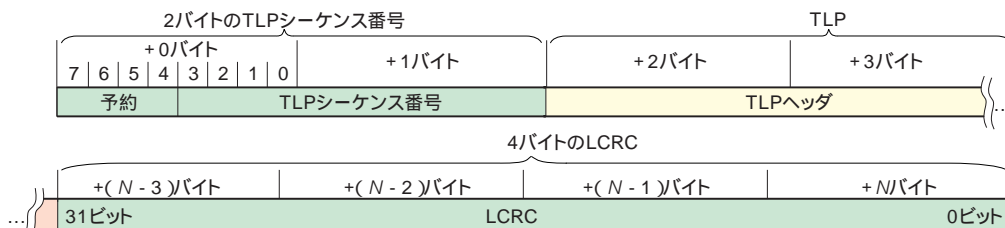



図8 DLLPの構造
データ・リンク層で生成するDLLPの構造を示す．DLLPは，ACK/NAKやフロー制御などのリンク管理に用いる．

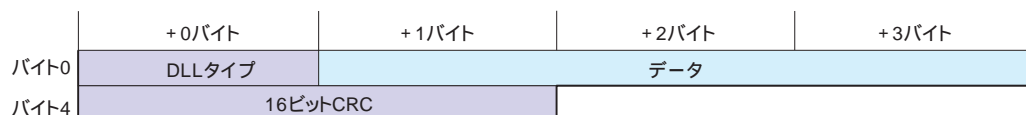



図9
PCI Express トランスミッタ回路ブロック図
各レーンへのバイト・データの配置を決めたあとの送信処理のブロック図を示す．スクランブルをかけたあと，8b/10bエンコードを行いシリアライズして差動で出力する．

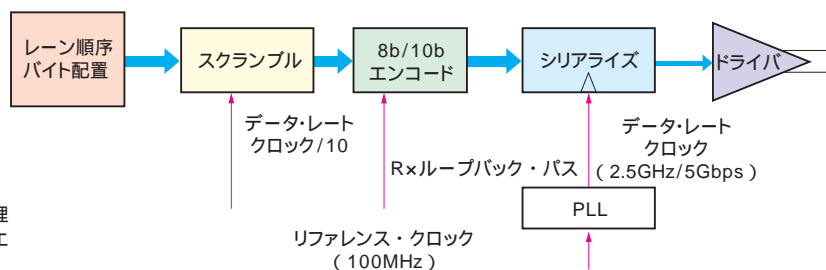
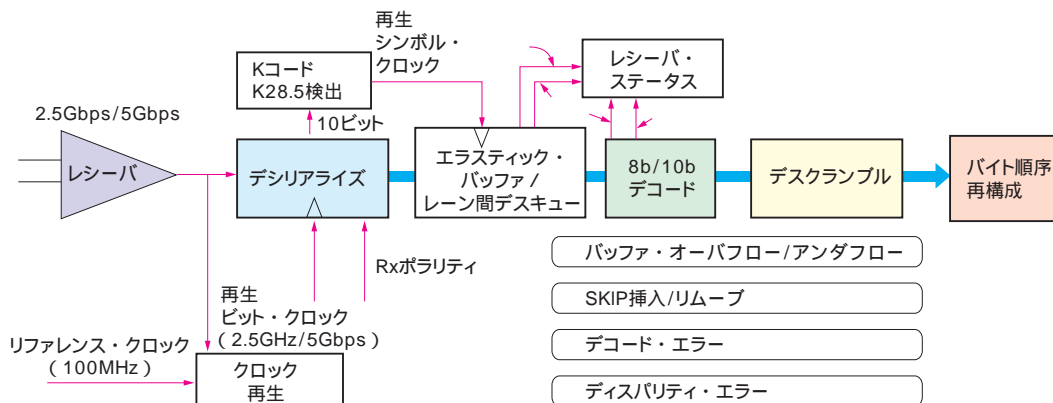



図10
PCI Express レシーバ回路ブロック図
各レーンの受信処理のブロック図を示す．送信処理とは逆に，デシリアライズしたあと，バッファなどを用いてレーン間のスキューを調整する．8b/10bデコードを行いデスクランブルしたあとに，バイト順に並べてデータを再生する．

制御に必要な情報を送信
- パワー・マネジメント関連

があります．

3）物理層

物理層は，データ・リンク層から受け取った情報をシリアル化し，リンクの受信側のデバイスと互換性を持つ周波数，帯域幅で送信します．物理層は，符号化やリンク制御を行う論理サブブロックと，実際の信号の送受信を行う電気サブブロックの二つに分けられます．

以下に論理サブブロックの役割を挙げます．
- インターフェースの初期化
- リンク幅およびレーン・マッピングのネゴシエーション
- リンク・パワー・マネジメント
- リセット/ホット・プラグ・コントロールとステータス

加えて電気サブブロックとして，以下の機能を備えます．
- フレーミング
- 8b/10bの符号化・復号化
- スクランブル/デスクランブル
- 多重化されたクロックとデータの再生
- シンボルの送信
- 受信側でのバッファリング
- レーン間デスキュー

電気サブブロックの詳細を次節で解説します．

2 物理層の技術

PCI Expressは，以下に挙げるような高速シリアル・インターフェースに共通の技術を使用しています．PCI Expressトランスミッタとレシーバの物理層の回路ブロックを**図9**と**図10**に示します．以下にそれぞれの回路ブロックの機能を解説します．

1）8b/10b符号化

GHzを超えるような高速信号においては，データ・パターンに依存して（高周波成分の）振幅が減少し，シンボル間干渉（ISI：Inter-symbol Interference）が生じます．シンボル間干渉を低減させるために，8b/10b符号化と後述のディエンファシスが併用されます．

8b/10b符号化は，DC（直流）成分の抑制とクロック・タイミング抽出を目的としています．8ビット・データを，‘1’または‘0’の連続を最大5サイクルに抑えた10ビット・パターンに変換します（**図11**）．10ビット・パターンには，データ256種類とパケットの先頭や終了などを示す制御符号（Kコード）16種類が含まれます（**表1**）．

短期間に信号変化が必ず含まれることで，受信側でクロックを再生しやすくなります．加えて，前に送り出されたパターンの`1'と`0'の数を見て，送り出しパターンの`1'と`0'を反転するディスパリティを行うことで，電位

図11　8b/10b符号化
8ビット(1バイト)のデータのうち，前3ビットと後5ビットを入れ替えて，それぞれ1ビットずつ加え，符号化する．8ビット・データから生成される256種類の10ビット・データに加え，Kコードと呼ばれる制御用の符号も利用する．

シンボル	データ	abcdei fghj出力	
		rd－	rd＋
D0.0	00h	100111 0100	011000 1011
D1.0	01h	011101 0100	100010 1011
D2.0	02h	101101 0100	010010 1011
D3.0	03h	110001 1011	110001 0100
…	…	…	…
K28.0	-	001111 0100	110000 1011
K28.3	-	001111 0011	110000 1100
K28.5	-	001111 1010	110000 1010

x1 レーン0　バイト0　バイト1　バイト2　バイト3
… a b c d e i f g h j …

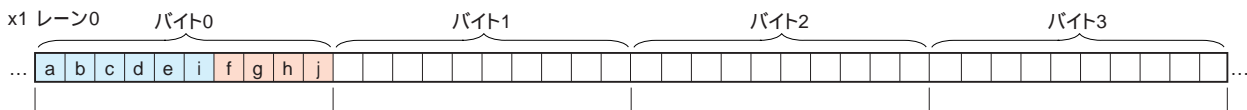
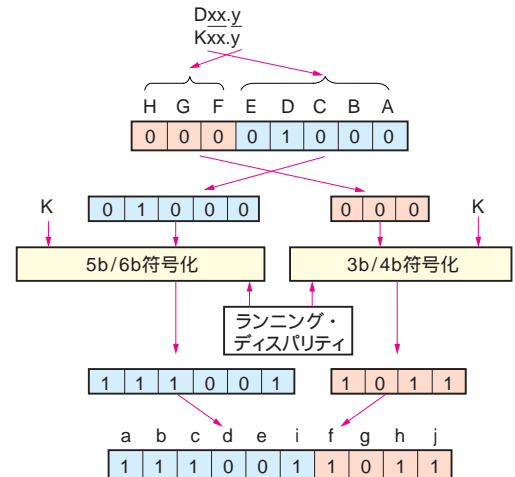



図12　ビット配置
1レーン(x1)の場合のビット配置を示す．複数レーンの場合もバイト・データのビット配置は同じである．

が一定になるようにバランスをとり，自動利得制御や情報の欠落を生じにくくします．ただし，8ビット・データを10ビット化するので，実際のデータ・レートは，物理層のデータ転送レートに対して20％小さくなります．

2）レーン間デスキュー

PCI Expressはレーン間のスキューに対する許容度が緩いので，レーン間で等長配線を行う必要がありません．

送信側では以下のスキューが許容されています．

2.5Gbps：500ps＋2UI(800ps)
5Gbps　：500ps＋4UI(1.6ns)

受信側では，より大きなスキューが許容されています．

2.5Gbps：20ns
5Gbps　：8ns

スキューは，受信側の各レーンに備えたFIFOで吸収します(デスキュー)．

3）フレーミング

データ・リンク層から引き渡されたパケット(DLLPやTLP)には，パケット先頭の目印として，KコードのSDP(Start of DLLP)かSTP(Start of TLP)が付加されます．また末尾には双方ともENDが付加されます．

フレーミングが行われた後，レーン構成に応じて各シンボルがレーンに割り当てられます．ビットは図12のように，ビットaで始まり，ビットjで終わるよう，一つのレーン上に配置されます．複数のレーンを持つリンクでは，最後のレーンまでバイト・データが順次配置された後，レーン0から同じように配置されます．

レーン数によっては，ENDの前にPAD(データ長を合わせるためのデータ)が挿入されることもあります．図13(a)には1レーンの，図13(b)には4レーンの場合のTLP

表1　PCI Expressで使用するKコード
例えば，受信側でのシンボルの切り出しはK28.5というKコードを基準にしている．そのため，一定間隔でK28.5が送信する必要がある．

Kコード	シンボル	名　前	内　容
K28.5	COM	Comma	レーン，リンクの初期化と管理に使用
K27.7	STP	Start TLP	TLPの始まりを示す
K28.2	SDP	Start DLLP	DLLPの始まりを示す
K29.7	END	End	TLPとDLLPの終わりを示す
K30.7	EDB	EnD Bad	送信途中でエラーが発生したTLPでENDの代わりに使用
K23.7	PAD	Pad	フレーミング時やリンク幅でのパディングとレーン順序のネゴシエーションで使用
K28.0	SKP	Skip	二つのポート間でのビット・レート差を補償するために使用
K28.1	FTS	Fast Training Sequence	LOsからLOに復帰するために使用(パワー・マネジメント)
K28.3	IDL	Idle	EIOS(Electrical Idle Ordered Set)で使用
K28.7	EIE	Electrical Idle Exit	EIEOS(Electrical Idle Exit Ordered Set)で使用(2.5Gbps以上のデータ・レート)

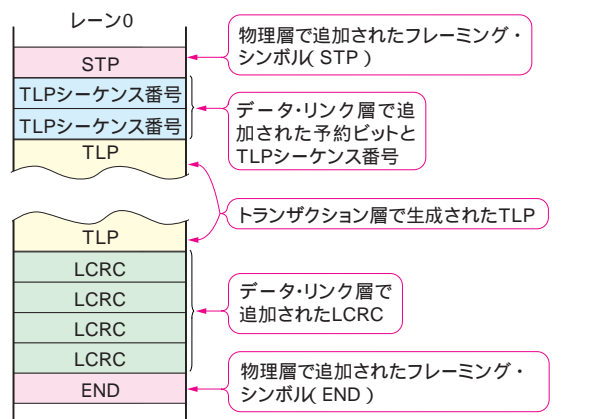
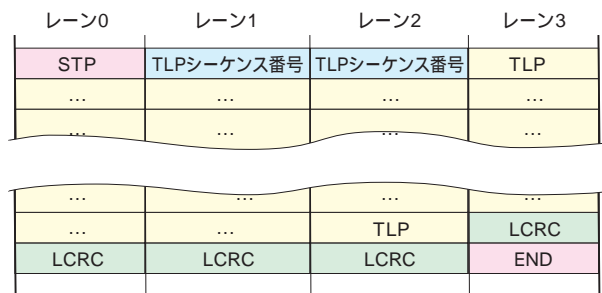



図13　TLPに対するフレーミング

（b）の場合は，レーン0→レーン1→レーン2→レーン3の順にバイト・データを割り当てる．場合によっては，PAD（データ長を合わせるためのデータ）を追加してレーン3にENDがくるように調整する．

図14　PCI Expressのレーン順序と極性の自由度

（b）のように，デバイス間で接続するレーンが異なっても，問題なく受信できる必要がある．同じように差動信号の極性が入れ替わっていても受信できる必要がある．これにより，基板設計が容易になる．CAPはAC結合用のキャパシタの意味．

のフレーミングを示します．

4）スクランブル/デスクランブル

同じデータが続いた際に，周波数軸上でエネルギーがある周波数成分に集中することを避けるため，8b/10b符号化の前にデータを線形フィードバック・シフト・レジスタ（LFSR）によりスクランブルします．LFSRの生成多項式$G(x)=X^{16}+X^5+X^4+X^3+1$です．受信側では，8b/10b復号化の後にデスクランブルします．

5）レーン順序と差動信号の極性

図14のように基板トレースの引き方を単純化できるよう，送信・受信間でレーン順序を変更できます．加えて差動信号の極性もレーンごとに反転でき，配線パターンの交差（ちょうネクタイ化）を防ぎます．

そのため，リンク確立時には，リンク幅，レーン割り当て，極性，そのほかを設定するために図15のようなTS1やTS2というトレーニング・シーケンスを発行します．

6）オーダード・セット

物理層の制御には，トレーニング・シーケンス以外にも，一連のKコードが挿入されるSKIP，EIOS（Electrical Idle Ordered Set），FTSといったオーダード・セットがあります．

7）コンプライアンス・パターン

PCI Expressでは，出力を直接測定器に終端すると起動時に物理層で検知し，図16のようなコンプライアンス・パターンを繰り返し自動発生する必要があります．これにより，電気的仕様のテストを容易にしています．

● 電気サブブロックで使われる基本的な技術

電気サブブロックは，実際にポート間を物理的に接続します．PCI Expressは以下に挙げるような高速シリアル・インターフェースに共通的な技術を使用しています．図17にPCI Expressの物理層回路ブロック図を，図18に物理層信号の仕様（Base Specification）を示します．

1）小振幅，差動伝送

PCI Expressの物理層は，ほかの高速シリアル・インターフェースと同じように，小振幅の信号レベルでかつ差動で伝送します．差動伝送には以下の特徴があります．

（1）コモン・モード・ノイズに対する耐性が高い（理論上で

図15 トレーニング・シーケンス例(TS2)

米国Tektronix社のTLA7000シリーズ・ロジック・アナライザのPCI Expressサポートを用いて，×16リンクから取り込んだデータを表示した．トレーニング・シーケンス中にレーン順序の判定と信号極性の判定を行っていることが分かる．

シンボル	K28.5−	D21.5	D28.5+	D10.2
現在のディスパリティ	0	1	1	0
ビット・パターン	0011111010	1010101010	1100000101	0101010101

図16 PCI Expressのコンプライアンス・パターン

コンプライアンス・パターンは，8b10b符号の中で一番低周波のパターン(5ビットの継続)と一番高周波のパターン('1'と'0'の1ビットごとの繰り返し)を含む4シンボル(合計40ビット)で構成されている．

図17 PCI Expressの物理層回路

伝送路も含めたデバイス間の1レーンの回路ブロックを示す．送信と受信を分離した1対1接続になっている．

は差をとるためゼロになる)．

(2) 信号レベルがシングルエンドの半分のレベルですむため消費電力を抑えられる．

(3) 互いに電磁界を打ち消し合うため，EMI(Electro-Magnetic Interference)を抑制できる．

ただし，差動信号間にスキュがあると，逆にEMIが増加したり，立ち上がり時間がなまったりします．そのため，差動トレース間のスキュはできる限り抑える必要があります．

またオフセット電圧がずれると，プラス方向(`0'→`1')とマイナス方向(`1'→`0')が対象でなくなったり，パル

図18
物理層信号の仕様(Base Specification Rev.1.1)
2.5Gbpsの場合のPCI Expressの送信側と受信側の電気的仕様を示す.

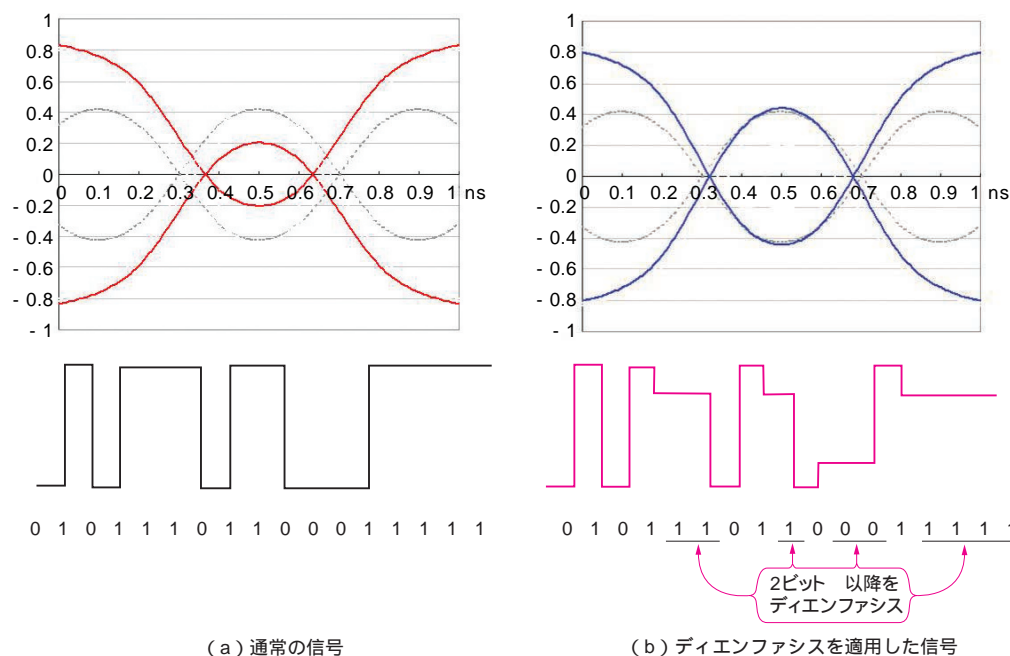



図19
ディエンファシス
高周波に対し損失がある伝送路においては,周波数が高いパターンは,(a)のように信号レベルが上がりきらない.そのため,パターンによってスレッショルド・レベルを横切るタイミングがずれ,ジッタを生じる.同じパターンが連続した場合(周波数が低い),(b)のように2ビット目以降の信号レベルを下げておくと,双方のパターンのレベル差が縮まり,ジッタが低減する.

スのデューティ比が変わったりします.ただしPCI ExpressではAC結合(伝送路に直列にコンデンサを入れてDC成分をカットする方式)のため,オフセット電圧は0Vです.

2) 1対1接続

一般的なバスの概念は,信号伝送路に複数のデバイスが接続されます.信号の高速化に伴い,信号の反射や劣化の原因となるスタブ(T分岐)を信号伝送路に設けないように,デバイス間を1対1で接続します.信号を分岐させる必要がある場合は,スイッチを経由します.

3) ディエンファシス

ディエンファシスとは,伝送路の高周波損失によって発生するシンボル間干渉を低減するために,同じビットが継続した場合,2番目以降のビットの振幅を下げることをいいます(**図19**).

継続した同じビットのことを非遷移ビットと呼びます.遷移ビットと非遷移ビットでは信号振幅が変わるので,それぞれ測定する必要があります.2.5Gbpsで−3.5dB,5Gbpsでは−3.5dBと−6dBが使用されます(±0.5dB).実際の波形を**図20**に示します.

4) クロック再生(CDR)

PCI Expressでは送信側で8b/10bのクロック・タイミングに合わせてデータを送信します.受信側ではクロック

コンプライアンス・テストは必要か？

コラム1

シリアル・インタフェースの標準規格団体は，年に数回，「プラグ・フェスタ（Plug Festa）」と呼ぶイベントを開催しています．プラグ・フェスタでは，現在開発中の機器を持ち込んで，すでに規格に適合した装置と接続してインターオペラビリティ（相互運用性）を確認する規格認証テスト（コンプライアンス・テスト）を受けられます．PCI-SIGではこのイベントのことを「コンプライアンス・ワークショップ（Compliance Workshop）」と称しています．年に数回，主に米国カリフォルニア州ミルピタス市で開催しています．

認証テストは，さまざまなデバイスと接続する必要があるPCI Expressデバイスのインタオペラビリティを保証するためにあります．インタオペラビリティが保証されると，インテグレーターズ・リストに掲載されます．

しかしPCI Expressデバイスとしての販売が目的ではなく，組み込み機器のようにクローズなシステムの内部で使うのであれば，インテグレーターズ・リストに掲載する必要はありません．わざわざ時間や費用をかけてまで認証テストを受ける必要はありません．

ただし，正しく動作させるために規格に適合した状態で使用していることを確認しておく必要はあると思われます．

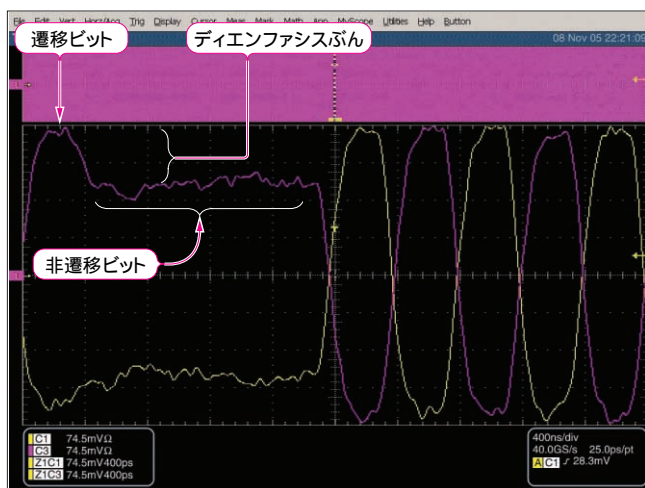



図20　ディエンファシスを適用した実際のPCI Expressの信号
2.5Gbpsの場合の－3.5dBのディエンファシスを行ったPCI Expressの信号．

再生（CDR：Clock Data Recovery）回路にて受信したデータの中からクロックを再生し，再生されたクロックでデータを取り出します．図21にCDR回路例を示します．

送信側と受信側の周波数差が±300ppm以下であれば，エラスティック・バッファがずれを吸収します．一方，EMI低減のためにSSC（Spread Spectrum Clocking）を使用する場合，リファレンス・クロック（100MHz）をシステム・ボードから供給します（コモン・クロック）．

外部クロック供給はRev.1.0aではオプションという形で位置づけられていました．Rev.1.1よりシステム・ボードのコンプライアンス・テストの必須項目となり，さらにRev.2.0ではPCI Expressの基本的な仕様書であるBase Specificationに含まれるようになりました．そのため，システム・ボードではリファレンス・クロックを供給できるようにしておく必要があります．

（a）PLLクロック・リカバリ方式

（b）位相インタポーレータ・クロック・リカバリ方式

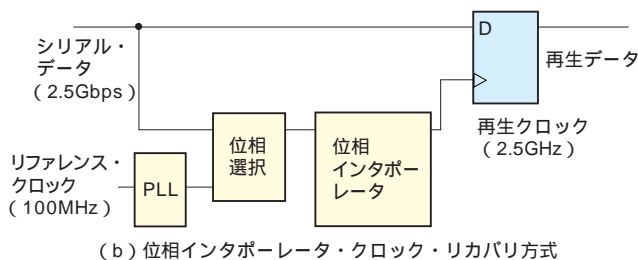
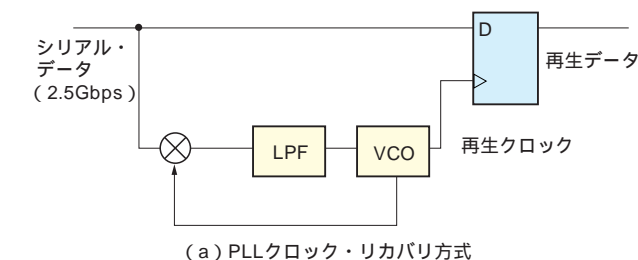



図21　クロック再生（CDR）回路
リファレンス・クロックを供給する場合，（b）を用いる．

参考・引用*文献

（1）* PCI-SIG；PCI Express Base Specification Revision 2.0，Dec ember 20，2006.
（2）野崎原生；PCI Expressデバイス＆システム設計の基礎の基礎，Design Wave Magazine，2006年1月号，pp.31-43，CQ出版社.
（3）碓井有三；高速伝送回路の基礎 システム機器設計者に必要なパラメータを理解する，ラムバス デベロッパ フォーラム ジャパン 2003 講演資料，2003年7月10日.

はたけやま・ひとし
日本テクトロニクス（株）

<筆者プロフィール>
畑山　仁．日本テクトロニクスにて営業技術統括部でシニア・テクノロジ・エクスパートとして従事．特にPCI Expressを中心に，高速ディジタル，高速シリアル・インターフェース分野を担当．